UNIVERSITÄT BERN

Faculty of Economics and Social Sciences

Department of Economics

Different Modeling Strategies for Discrete Choice Models of Female Labour Supply: Estimates for Switzerland

Reto Nyffeler

05-08

July 2005

DISCUSSION PAPERS

Gesellschaftsstrasse 49
CH-3012 Bern, Switzerland
http://www.vwi.unibe.ch

Different Modeling Strategies for Discrete Choice Models of Female Labour Supply: Estimates for Switzerland

Reto Nyffeler*
University of Bern

July 2005

Abstract

In recent applications of discrete choice models of labour supply considerable attention has been devoted to strategies to increase the flexibility of models for a better fit to the data. These include the introduction of random parameters, fixed cost of work or flexible functional forms of preferences. Based on estimates of models of recent studies this paper compares these different modeling strategies. Results for Swiss data show that the traditional way to interpret fixed cost of work is ad hoc. Furthermore our results indicate that care should be taken when using very general function forms of preferences.

JEL Classification C25, C52, H31, J22

Keywords: multinomial logit, household labour supply, taxation, microsimulation

Correspondence:
Reto Nyffeler
Department of Economics
University of Bern
Gesellschaftsstr. 49
3012 Bern
Switzerland
reto.nyffeler@vwi.unibe.ch

*I am grateful above all to Michael Gerfin for his help and advice. I would also like to thank Robert Leu, Christian Rütschi and the participants of IZA summer school 2005 for useful comments. Some of the models are estimated with Biogeme which is provided for free by Michel Bierlaire.

1 Introduction

Over the last decade the discrete choice approach of labour supply analytics gained more and more popularity in assessing the impact of public policies on work incentives. The approach can easily handle non-linear and possibly non-convex budget sets caused by taxation. In addition it avoids the MaCurdy, Green, and Paarsch (1990) critique that coherency of the model implicitly limits the range of elasticities that can be obtained.

Recent studies of discrete labour supply focused very much on improving the models capability in explaining the peaks in the hours distribution. This was done by gaining flexibility through the introduction of random parameters, state specific constants or fixed cost of work (see for example (Van Soest 1995)). More recently flexibility of the models was further increased by allowing for broader functional forms of preferences. Discrete choice analytics does not need explicit expressions for both the direct utility function and the labour supply function (or the indirect utility or expenditure function) and therefore very general functional forms of preferences are principally possible. The move to more general preferences is also motivated by the repeated rejection of restrictions usually imposed on household preferences in the standard approach

The present paper compares different discrete modeling strategies used in recent studies and tries to assess the value of the various model extensions made to gain flexibility. We start with a basic structural model with a quadratic direct utility function and then extend it by allowing for fixed costs of work and unobserved individual preference heterogeneity. This model similar to the one of Van Soest (1995) or Blundell, Duncan, McCrae, and Meghir (2000) can be seen as the standard model used in policy analysis. We then move on to models with more flexible functional forms of preferences. These flexible models are taken from two recent contributions, one by Van Soest, Das, and Gong (2002) and the other by Bargain (2004). In the framework of Van Soest, Das and Gong the direct utility function is approximated by a nonparametric series approximation in hours and income. In this way they introduce a structural nonparametric labour supply model which can be used for all sorts of policy analysis. Bargain suggests two generalizations of the structural model that are more radical. In the first suggestion preference parameters are allowed to be alternative specific, that is utility can depend on disposable income in a fully flexible way over working hours. The second generalization allows the utility of each alternative to depend on disposable income as well as on wage rates and non-labour income. In some sense then preferences are price and income dependent and do not verify the restrictions normally imposed on household preferences in the standard approach.

The models are estimated with data from the Swiss Income and Expenditure Survey 1998 (SIES) using simulated maximum likelihood. The model evaluation is mainly based on three criteria: explanatory power, usability in simulation exercises and economic foundation. Based on the estimates of the models that perform best with respect to these criteria we simulate the female labour supply effects of the introduction of a individual tax system in Switzerland.

The paper is structured as follows: Section 2 gives an overview of the random utility models (RUMs) used in this paper to represent utility maximizing discrete choices. Section 3 introduces the different modeling strategies and model specifications. In section 4 the empirical results concerning the performance of the various models and the simulation exercises are presented and discussed. Section 5 concludes.

2 Logit Models for Multiple Choices

In discrete choice labour supply modeling the labour supply decision is described as the utility maximizing choice between discrete hours alternatives. A prominent way to model utility maximizing discrete choices are random utility models (RUMs). They are the basis of all the choice models used in this paper and can be derived as following.[1] A decision maker i faces a choice among J alternatives. Each alternative provides a certain level of utility. From alternative J the decision maker obtains utility U_{ij}, $j = 1, ..., J$. Alternative j is chosen if $U_{ij} > U_{ik}$ for all $k \neq j$. The decision maker's utility can be decomposed as

$$U_{ij} = V_{ij} + \varepsilon_{ij}, \tag{1}$$

where V_{ij} is a function which relates observed factors to the decision maker's utility. These factors are attributes of the alternatives, X_{ij} $\forall j$, and some attributes of the decision maker, S_i. V_{ij} depends on unknown parameters β_j which have to be estimated. The function is denoted $V_{ij} = V(X_{ij}; S_i, \beta_j)$ $\forall j$ and is called representative utility. Factors that are not included in V_{ij} but affect utility are captured by ε_{ij}. This part of the utility is unknown and assumed to be random. It can be seen as the error made in evaluating alternative j. Since ε_{ij} is simply the difference between U_{ij} and V_{ij} this decomposition is completely general.

The logit model is obtained by assuming that each ε_{ij} is independently, identically distributed extreme value. The density and cumulative distribution of ε_{ij} are respectively

$$f(\varepsilon_{ij}) = e^{-\varepsilon_{ij}} e^{-e^{\varepsilon_{ij}}}$$

and

$$F(\varepsilon_{ij}) = e^{-e^{-\varepsilon_{ij}}}.$$

Mc Fadden (1974) has proved that under this assumption the probability that decision maker i chooses alternative k is

$$\begin{aligned} P_{ij} &= Prob(V_{ik} + \varepsilon_{ik} > V_{ij} + \varepsilon_{ij} \; \forall k \neq j) \\ &= \frac{e^{V_{ij}}}{\sum_k e^{V_{ik}}}. \end{aligned} \tag{2}$$

[1] See Train (2003) for an excellent overview.

Representative utility can either be specified to be linear or nonlinear in parameters. If parameters enter representative utility nonlinearly estimation is more difficult because the log-likelihood function may not be globally concave.

Since the logit probabilities take a closed form, the traditional maximum-likelihood procedures can be applied. The log-likelihood function is given by

$$LnL(\beta_j) = \sum_{i=1}^{I} \sum_{i} d_{ij} \; lnP_{ij}, \tag{3}$$

where $d_{ij} = 1$ if person i chose j and zero otherwise. The simplicity of the logit model is a strong advantage. But logit models have some clear limitations. They can only represent systematic taste variation but not random taste variation and they imply proportional substitution across alternatives, that is logit models exhibit the IIA property. One model that obviates these disadvantages is the mixed logit model (Brownstone and Train 1998, McFadden and Train 2000).

The mixed logit choice probability can be derived in several ways from utility maximizing behavior. The following derivation is based on the random coefficient interpretation (Revelt and Train 1998).[2] The utility of person i from alternativ j is given by

$$U_{ij} = V_{ij} + \varepsilon_{ij} = V(X_{ij}; S_i, \beta_i) + \varepsilon_{ij}, \tag{4}$$

where X_{ij}, S_i and ε_{ij} are defined as before and β_i is a vector of coefficients for person i.[3] If utility is linear in β_i and we abstract from S_i utility can be written as $U_{ij} = \beta_i' x_{ij} + \varepsilon_{ij}$. This specification is the same as for logit, except that now the coefficients β_i vary randomly over the decision maker rather than being fixed. The coefficient vector for each decision maker can be expressed as the sum of the mean, b, and individual deviation, η_i. Utility is then $U_{ij} = b' x_{ij} + \eta_i' x_{ij} + \varepsilon_{ij}$. The unobserved portion of utility is $\eta_i' x_{ij} + \varepsilon_{ij}$. This term is correlated over alternatives due to the common η_i. Because of this correlation, mixed logit does not exhibit the independence from irrelevant alternatives property.[4] If we knew the decision maker's taste, that is, if we knew the value of β_i the conditional choice probability would be standard logit since ε_{ij}'s are iid extreme value, that is

$$L_{ij}(\beta_i) = \frac{e^{V_{ij}(\beta_i)}}{\sum_{k=1}^{J} V_{ik}(\beta_i)}.$$

[2]An other popular interpretation is based on error components. But since here the stress is more on individual taste variation and less on substitution patterns the random coefficient interpretation seems more natural.

[3]For notational simplicity we use here β_i instead of β_{ij}. However it is no problem to generalize mixed logit to allow for alternative specific random coefficients. In order to avoid the IAA property either the variance of these random coefficients has to be the same for all alternatives or the random coefficients are allowed to be correlated over alternatives.

[4]Mixed logit allow for very general patterns of correlation and hence very general patterns of substitution. McFadden and Train (2000) have shown that any random utility model can be approximated by mixed logit.

Since β_i is not given (we can estimate b but can not observe η_i for each decision maker), the (unconditional) choice probability is this logit formula integrated over all values of β_i

$$P_{ij} = \int \frac{e^{V_{ij}(\beta)}}{\sum_{k=1}^{J} V_{ik}(\beta)} f(\beta) d\beta. \tag{5}$$

Models of this form are called mixed logit because the choice probability is a mixture of logits with $f(\beta)$ as the mixing distribution.[5] The mixing distribution may be discrete or continuous. In the discrete case the mixed logit becomes the so called latent class model.[6] As most applications of mixed logit we assume the density of β to be continuous and more specific to be normal with mean b and covariance W. In this case the choice probability is given by

$$P_{ij} = \int \frac{e^{V_{ij}(\beta)}}{\sum_k e^{V_{ik}(\beta)}} \phi(\beta|b, W) d\beta, \tag{6}$$

where $\phi(\beta|b, W)$ is the normal density with mean b and covariance W. The parameters to be estimated are those of the mixing distribution $f(\beta)$, b and W.

Since there is no closed form expression for the choice probabilities in mixed logit we approximate the probabilities by simulation and maximize the simulated log-likelihood function. In particular for given b and W a value of β is drawn from $f(\beta|b, W)$. This value is labeled β^r with the superscript r=1 referring to the first draw. Using this draw the standard logit formula $L_{ij}(\beta^r)$ is calculated. This process is repeated for many draws and the results are averaged. This average is the simulated probability:

$$\check{P}_{ij} = \frac{1}{R} \sum_{r=1}^{R} L_{ij}(\beta^r), \tag{7}$$

where R is the number of draws. $\check{P}_{ij}$ is an unbiased estimator of P_{ij}. Its variance decreases as R increases. The simulated probabilities are inserted into the log-likelihood function to give a simulated log-likelihood:

$$SSL = \sum_{i=1}^{I} \sum_{j=1}^{J} d_{ij} ln \check{P}_{ij}, \tag{8}$$

where d is defined as above. The maximum simulated likelihood estimator is the value of b and W that maximizes the simulated log likelihood.

[5]The standard logit model is a special case where $f(\beta)$ is degenerate at fixed parameters b: $f(\beta) = 1$ for $\beta = b$ and 0 for $\beta \neq b$.

[6]In the latent class model $\beta = b_m$ with probability s_M. The choice probability for this model is given by

$$P_{ij} = \sum_{m=1}^{M} s_m \frac{e^{V_{ij}(b_m)}}{\sum_k e^{V_{ik}(b_m)}}.$$

Here it assumed that the population consists of M segments each with its own choice behavior or preferences. The share of segment m in the population, s_m, is estimated along with the b_m's for each segment (see Bargain (2004) for a recent application to labour supply estimation.

3 Specification of the Models

3.1 Structural Models of Labour Supply

We start with describing a static neo-classical structural labour supply model for single decision makers. Following Keane and Moffitt (1998), Blundell, Duncan, McCrae, and Meghir (2000) and (Van Soest, Das, and Gong 2002) we assume that the answer to the desired hours question is based upon maximizing

$$U_{ij} = V(Y_{ij}, H_j; S_i, \beta) + \varepsilon_{ij}, \tag{9}$$

where Y_{ij} is net household income, H_j are female hours of work and S_i are household characteristics. The net household income Y_{ij} is given by

$$Y_{ij} = w_i H_j + Y_{im} + Y_{inl} - T(w_i H_j, Y_{im}, Y_{inl}; S_i),$$

where w_i is the female's wage rate, Y_{im} the husband's labour income, Y_{inl} is the household's non labour income and $T(w_i H_j, Y_{im}, Y_{inl}; S_i)$ are the tax payments. As in Blundell, Duncan, McCrae, and Meghir (2000) the utility function is specified to be quadratic and is given by

$$U_{ij} = \beta^{YY} Y_{ij}^2 + \beta^{HH} H_j^2 + \beta^{YH} Y_{ij} H_j + \beta^{Y} Y_{ij} + \beta^{H} H_j \quad \text{for } j = 1, ...J. \tag{10}$$

Observed heterogeneity is introduced by assuming that

$$\beta^H \quad = \quad \beta_{h0} + \beta_h' S_i. \tag{11}$$

In principle there is no theoretical reason to only allow β^H to vary with S. However the identification of the effects of S via different β's is often difficult. In addition β^H is an attractive choice for interpreting the results. It implies that the marginal utility of work varies linearly with S. The sign of the β_h coefficients directly determines if the variables in S have a positive (positive sign) or a negative (negative sign) effect on the marginal utility of work. This basic model will hereafter be referred to as model S1. It has two major shortcomings. First, it does not fit the data, in the sense that it underpredicts nonparticipation and overpredicts part-time jobs involving a few hours a week. Second, it does not allow for unobserved individual heterogeneity. Several methods have been used to overcome the first shortcoming. Van Soest (1995) introduced some hours specific constants on an ad hoc basis in the utility function. These may reflect costs of finding a part-time job. An alternative with a more attractive economic interpretation is the incorporation of fixed costs of work (Callan and Soest 1996).[7] Fixed costs are the costs an individual has to pay to get to work. By subtracting them from income for the strictly positive working hours they can be introduced into the model in a natural way. For countries with very high costs of childcare like Switzerland fixed costs of work are mainly made up

[7] Another alternative would be the approach of Dickens and Lundberg (1993), who incorporate demand-side restrictions on hours worked explicitly, but this model requires strong assumptions for identification.

by childcare costs. In principal it would therefore be preferable to proceed as Blundell, Duncan, McCrae, and Meghir (2000) and use sample information on hourly prices of childcare to account for childcare expenditures. However data on childcare costs for Switzerland are of a too poor quality. We therefore have to define a fixed costs equation in terms of a set of observable variables. Fixed costs are assumed to be not stochastic and are specified as

$$F_i = \delta' Z_i, \tag{12}$$

where Z_i is a subset of S_i.[8] For all states $j > 0$ utility expression 9 becomes

$$U_{ij} = V(Y_{ij} - F_i, H_j; S_i, \beta) + \varepsilon_{ij}. \tag{13}$$

If utility increases with income, fixed costs decrease the utility of working compared to the utility of not working, thereby increasing the probability of nonparticipation. This model will be referred to as model S2. The second shortcoming can be removed by adding an error term to one of the parameters of the utility function. We follow this strategy and assume that unobserved heterogeneity enters through the parameter β^Y:[9]

$$\beta_i^Y \quad = \quad \beta_{y0} + v_{iy}, \tag{14}$$

where $v_{iy} \sim N(0, \sigma^2_{v_{iy}})$. The model with fixed costs of work and random preferences will be referred to as model S3.

In contrast to the continuous labour supply model imposing quasi-concavity of preferences a priori is not necessary for coherency of the model in discrete choice analysis (Van Soest, Kapteyn, and Kooreman 1993). Quasi-concavity of the utility function can be checked ex post. In this way the MaCurdy critique that elasticities are largely determined a priori (through the quasi-concavity restriction) can be avoided. In fact since only utility in the finite choice sets matters quasi-concavity is not even necessary for economic interpretation of the model. The only restriction required for economic interpretation is that utility is increasing with income. This restriction we need since we assume that everyone always chooses a point on the frontier of the budget set rather than in the interior. However we will not impose this condition a priori before estimation but check it ex post (Van Soest, Das, and Gong 2002).[10] These considerations are also valid for the more flexible models which follow in the next sections.

[8] We also experimented with stochastic fixed costs. However this did not help to improve the fit of the model.

[9] This choice is driven by the fact that the structural models must be comparable with the more flexible models where only the income terms remain in the utility function. However the model was also estimated with unobserved heterogeneity entering through β^H. The fit of the model did not improve at all.

[10] Another restriction that can be checked ex post is the monotonicity restriction with respect to labour supply. Interestingly the recent literature seems not very concerned about this restriction (see for example Bargain (2004) or Van Soest, Das, and Gong (2002)).

3.2 More Flexible Models of Labour Supply

The improvements of model S1 through the introduction of fixed costs and random coefficients are strategies to increase flexibility of structural models for a better fit to the data. However the specification of preferences in these models is still restrictive. Discrete choice labour supply would allow for more flexible specifications since in contrast to continuous models discrete choice analytics does not need explicit expressions for both the direct utility function and the labour supply function (or the indirect utility or expenditure function).[11] The question is therefore if it is reasonable and possible to use more flexible specifications of preferences. We consider two recent contributions put forth by Bargain (2004) and Van Soest, Das, and Gong (2002) respectively which propose more flexible models of labour supply.

3.2.1 A Structural Labour Supply Model with Nonparametric Preferences

Van Soest, Das, and Gong (2002) introduce a structural nonparametric labour supply model (hereafter SNP model). Basically they replace the direct utility function of the structural model S3 with a flexible polynomial expansion. In this way they maintain the economic structure of model S3 (utility maximization under a complex budget set) and combine it with a nonparametric specification of the utility function. Since the general structure of this flexible model remains unchanged it can be represented by expression 13:

$$U_{ij} = V(Y_{ij} - F_i, H_j; S_i, \beta) + \varepsilon_{ij}.$$

The direct utility function is specified as higher order polynomial in its arguments H and Y:

$$U_{ij} = \sum_{p=0}^{K} \sum_{q=0}^{K-p} \beta^{Y^p H^q} H^p (Y - F)^q. \tag{15}$$

K is the order of the polynomial and determines the flexibility of the utility function. Since for K equal to two we get the model discussed in the previous section K has to be larger than two. If K is allowed to be arbitrarily large, U_{ij} is able to approximate any utility function in a given compact set of relevant hours income combinations. However for finite sample size the order of the polynomial that can be used is limited.[12] In the empirical section we will consider the case of K=5 the largest value of K Van Soest, Das and Gong used in their work. As in the previous section observed and unobserved heterogeneity is assumed to enter through the parameters β^Y and β^H

$$\begin{aligned} \beta_i^Y &= \beta_{y0} + v_{iy} \\ \beta^H &= \beta_{h0} + \beta_h' S_i, \end{aligned}$$

[11] See for example Creedy and Duncan (2002).

[12] Asymptotics requires that K tends to infinity much slower than the number of observations.

where $v_{iy} \sim N(0, \sigma^2_{v_{iy}})$. Fixed costs are specified and introduced into the model as before. An important comment can be made concerning the identification of fixed costs. The identification of fixed costs can be intuitively explained by the lack of observations with low working hours. In the case of a fully nonparametric utility function it could be that the utility function itself could pick up the gap in the distribution at low hours, by assigning lower utility to such hours values. Fixed costs would then be nonparametrically unidentified (Van Soest, Das, and Gong 2002). However since S enters the utility function and fixed costs in a restrictive way this should not be a matter of concern here.

3.2.2 Unconstrained and Non-standard Models of Labour Supply

Bargain (2004) suggests two generalizations of model S3 which relax the restrictions on household preferences imposed in this model step by step. His generalizations are more radical than the one from Van Soest, Das, and Gong (2002). However despite their generality both models maintain a utility maximizing interpretation.

Unconstrained Model (Model U) In this model preference parameters are alternative specific, that is utility can depend on consumption in a fully flexible way across working hours. A direct interpretation of any of the parameters is no more possible. The model which nests the structural models from section 3.1 is given by

$$U_{ij} = V(Y_{ij}; S_i, \beta_j) + \varepsilon_{ij}, \tag{16}$$

where Y_i is given as before.[13] Using the quadratic form the utility function for this model has the form

$$U_{ij} = \beta_j^{YY} Y_{ij}^2 + \beta_j^Y Y_{ij} + \delta_j \quad \text{for } j = 1, ...J. \tag{17}$$

Observed and unobserved heterogeneity is written as

$$\begin{aligned} \beta_j^Y &= \beta_{y0j} + \beta'_{yj} S_i + v \qquad (18) \\ \delta_j &= \delta_{0j} + \delta'_j S_i + \sum_{k=1}^{L} \sum_{l=k}^{L} \delta_j^{kl} s_i^k s_i^l. \end{aligned}$$

Only $J-1$ sets of parameters δ_j can be identified. For the first alternative the δ coefficients are therefore set to zero. Since disposable income is alternative specific all J β coefficients can be estimated.

Non-standard or General Model (Model G) So far wage rates and non-labour income influence labour choices only through disposable income. This is consistent with the standard or unitary approach and implies income pooling

[13]See the appendix of Bargain (2004) for the identification of the constraints imposed by model S3 on the unconstrained model

and common preferences within a household. Bargain's second generalization allows each alternative to depend on disposable income as well as on wage rates and non-labour income. In some sense then preferences are price- and income-dependent. Such a non-standard model could have the following form:

$$U_{ij} = V(Y_{ij}, w_i, Y_{im}, Y_{inl}; S_i, \beta_j) + \varepsilon_{ij}, \tag{19}$$

where the variable definitions remain the same as above. Utility of alternative j is now not only dependent on disposable income of the household but also on female's wage rate, the labour income of the husband and non-labour income. This model can be rationalized in different ways. It can be related to the collective approach (Chiappori 1988), it can be made consistent with the life cycle framework and finally the model could reflect constraints from the demand side.[14] However the model does not allow to discriminate between these different approaches. Keeping the quadratic form the utility function of the model could be specified as

$$\begin{aligned} U_{ij} \;=\; & \beta_j^{YY} Y_{ij}^2 + \beta_j^{Y} Y_{ij} + \beta_j^{ww} w_i^2 + \beta_j^{Y_m Y_m} Y_{im}^2 + \beta_j^{Y_{nl} Y_{nl}} Y_{inl}^2 \\ & + \beta_j^{w Y_m} w_i Y_{im} + \beta_j^{w Y_m} w_i Y_{inl} + \beta_j^{Y_m Y_{nl}} Y_{im} Y_{inl} + \beta_j^{w} w_i + \beta_j^{Y_m} Y_{im} \\ & + \beta_j^{Y_{nl}} Y_{inl} + \beta_j^{wY} w_i Y_{ij} + \beta_j^{Y_m Y} Y_{im} Y_{ij} + \beta_j^{Y_{nl} Y} Y_{inl} Y_{ij} + \delta_{ij} \\ & \text{for } j = 1, ...J \end{aligned} \tag{20}$$

with:

$$\begin{aligned} \beta_j^Y &= \beta_{y0j} + \beta_{yj}' S_i + v \\ \beta_j^R &= \beta_{r0j} + \beta_{rj}' S_i \quad \text{for} \quad r = w, Y_m, Y_{nl} \\ \delta_{ij} &= \delta_{0j} + \delta_j^{'} S_i + \sum_{k=1}^{L} \sum_{l=k}^{L} \delta_j^{kl} s_i^k s_i^l. \end{aligned} \tag{21}$$

4 Data and Empirical Results

4.1 Data

The data used in this analysis are drawn from the Swiss Income and Expenditure Survey 1998 (SIES). Over 9000 households participated in this survey conducted by the Swiss Federal Office of Statistics. The survey is primarily used for the periodical revisions of the Swiss National Consumer Price Index. Besides the detailed expenditure data including tax and social security payments the survey also provides information about all sources of income as well as about labour supply of each household member.

For our estimations and simulation exercises we need net incomes and hence taxes for all hours alternatives. Unfortunately the tax system in Switzerland is

[14] Bargain (2004) provides a short explanation how the model could be made consistent with the different approaches.

rather complicated. The majority of taxes consists of cantonal and communal taxes that vary considerably across cantons and communities. In other words there are 26 different tax systems. Matters get further complicated by the fact that within these 26 tax systems the communal tax rates vary as well. Given the complexity of the tax system we use a simplified tax model. Instead of communal tax factors we apply the tax factor of the canton's capital to everyone living in the respective canton. Cantonal and federal taxes are computed according to the published tax tables. Communal taxes are the capital's tax factor times cantonal taxes.

For the empirical analysis we select married or de-facto couples, aged between 20 and 65, who are employed or voluntarily unemployed. Students, self employed, unemployed or retired people are excluded from the sample. Moreover people who work more than 60 hours a week, households with more than four children or with more than two decision makers are selected out. People with very high levels of non-labour income and individuals with wages below or above the 1st and 99th percentiles of the wage distribution were also discarded. Since men's participation rate is very high (99.8%) and almost all men work full time the empirical analysis fully concentrates on female labour supply. Working hours of men are fixed at the observed value. With this selection the sample contains 2305 households. Table 1 displays descriptive statistics for the sample. Since gross wage rates for non-working individuals are not observed these wages are predicted using the standard Heckman two-step estimation procedure. For workers the actual wage rates are used.[15] Figure 1 displays the distributions of predicted and observed wage rates for part-time and full-time female workers respectively. The fit is more satisfying for full time workers than for part time workers. In both cases the predicted wage distribution is more concentrated around the mode. Figure 2 shows the distribution of female weekly working hours. A significant portion of females in couples does not participate in the labour market and the fraction of part-time working females is quite large. For the empirical analysis we assume that women have the following discrete choice set: H$\in \{0, 8, 16, 25, 33, 42\}$.

4.2 Empirical Results

Table 2 displays the estimation results for the structural models S1, S2 and S3. The interpretation of the parameters has to be made with caution. Directly interpretable are the interactions between hours worked and household characteristics. These coefficients determine how marginal utility changes with household characteristics. Age and the presence of children decrease marginal utility of work. In the case of children the effect is stronger for preschool children than for schoolaged children. High education increases the marginal utility

[15]We are aware of the fact that this approach in principle does not lead to consistent estimators since it assumes that wage rates of nonworkers are predicted without errors. For consistent estimators it would be necessary to take the wage rate prediction errors explicitly into account for example by integrating out the disturbance term of the wage equation in the likelihood (Van Soest 1995).

of work. These results seem consistent with intuition and are in line with other studies (see for example Duncan and Harris (2002).

Fixed costs of work significantly decrease with the number of children. This counterintuitive result was also found by other studies (see for example Duncan and Harris (2002) and Van Soest, Das, and Gong (2002)). Combined with the effect of children on preferences, this finding could mean that for women with children, working a small number of hours per week is particularly attractive (see Van Soest, Das, and Gong (2002)). On the whole the fixed cost coefficients are implausible high. Depending on the model average fixed costs represented by the constant term are more than 100% of the average earnings of working women. Since fixed costs are constant over hours alternatives utility of working is particularly decreased for low working hours.[16] This is reflected by the kink in the estimated indifference curves in figure 3 at 8 hours of work per week.[17] What exactly this coefficient measures is not clear. One possibility is that the coefficient captures demand side aspects like insufficient availability of low hours jobs. Considering these results the view that fixed costs are a structural way to increase the flexibility of a model can be questioned. The interpretation of the estimated coefficients as fixed costs of work seems rather ad hoc and not very plausible. As an alternative to the explicit inclusion of fixed costs we could allow the taste shifters S to enter utility in a less restrictive way. The utility function would then represent preferences in which fixed costs are already captured. Another possibility would be to collect data about fixed costs of work. For Switzerland this would mean to collect data about hourly prices of child care and child care usage and then proceed as Blundell, Duncan, McCrae, and Meghir (2000).

The inclusion of random preferences seems to considerably improve the precision of the preference parameters and leads to quite a large increase of the hours and fixed cost coefficients. Given the significantly estimated standard deviation of the distribution of the random coefficient there seems to exist considerable heterogeneity concerning income preferences.

The considerations just made about the interpretation of the parameters, in particular what we said about implausible high fixed cost coefficients and demand side contamination, remain also valid for the SNP model. The estimation results for this model are displayed in table 3. Due to convergence problems this model is estimated without individual heterogeneity. Given the large number of parameters and the difficulty of interpretation the estimation results for the U and G model are omitted.[18] These models have been estimated with and without random parameter. However in a likelihood ratio test the conditional logit model (the model without unobserved preference heterogeneity) could not be rejected. In the following comparison of the models we thus ignore individual heterogeneity for the models SNP, U and G.

[16] We also tried to estimate a model with alternative specific fixed cost coefficients. However the model was not identified in this case.

[17] The indifference curves are plotted for a benchmark individual: age 45, no children, medium education level.

[18] Results available upon request from the author.

Table 4 contains some information about the fit of the estimated models. In the upper part of the table the observed frequencies are compared with its average estimated value over all households. Not surprisingly the flexible models SNP, U and G perform best in this respect. Interestingly predicted probabilities of model S3 seem not to be much more accurate than those of model S1. The relatively good performance of model S1 may be due to the fact that in Switzerland female part-time work is widely spread and the pattern of working hours is not as rigid as in other countries. Another measure of fit displayed in table 4 is the pseudo-R2 or Likelihood Ratio Index of McFadden (1973). The measure is defined as $1 - LogL_e/LogL_0$, where $LogL_e$ is the log-likelihood function for the estimated model and $LogL_0$ is log-likelihood function when all parameters are set to zero. The definition of the measure implies that it is always between zero and one. According to this measure the general model clearly provides a better fit than the standard models and the flexible standard models dominate the simple structural model S1. Again this comes with no surprise.

Table 5 displays the log-likelihood values for the models S3, SNP, U and G. In addition it provides the LR statistics and the relevant critical values at the 1% significance level. Tests of model S3 against model SNP and U result in a rejection of model S3 in both cases.[19] This implies that from a statistical point of view the restrictions in the structural model S3 are too restrictive. Furthermore line three of table 5 shows that in a test of the U model versus the G model the standard model is rejected. In a static setting, assuming partial equilibrium, this is nothing else than a test of the unitary model against a model with price-dependent preferences that does not verify Slutsky conditions or pooling (Pollak 1977). Thus this is strong evidence against the unitary approach. So far we agree with Bargain (2004).

As stated in section 3.1 the only coherency restriction we really need for the economic interpretation of the models is that utility is monotonically increasing in income. This restriction is satisfied for all observations and labour supply choices in the models S1, S2, S3 and SNP. In the models U and G however marginal utility of income is positive for only 87.3% and 52.9% of the labour supply choices respectively. Thus the increased flexibility of model U and G has the advantage of capturing broader preference heterogeneity but has the disadvantage that a significant portion of the observations behave in contradiction to economic theory. For model U the problem of the violation of the coherency restriction can be solved by a restricted estimation. Practically this can be done by penalizing the log-likelihood for observations at which utility of a corresponding interior point of the budget set exceeds utility of the point on the edge. For model G this procedure seems not to make sense since the percentage of observations with positive marginal utility of income is too low. Re-estimating model U imposing the monotonicity restriction leads to a slightly higher likelihood value than in the unrestricted model. The log-likelihood value and the corresponding likelihood ratio are displayed on line four of table 5.

[19]Model SNP also yields a clearly higher AIC value than model S3.

Despite the possibility of a restricted estimation we only use model S3 and SNP for the simulation exercises. This is for several reasons: First, it is true that economic theory does not impose the functional form of the utility function but it is not obvious why the effect of income should differ over the supply choices as in model U and G. Second, it is not clear why these different effects should remain constant after a tax reform. Third, the number of parameters in model G and U is very large and some of these are imprecisely estimated. Fourth, the parameterization depends on the chosen discretization of the choice set. The more discrete hours level are used the more parameters you get. Overall model SNP seems to perform best. The model is flexible but has still a rich economic structure, it fits the data well and all supply choices in our sample exhibit positive marginal utility of income. Model S3 which can be seen as the standard model in policy analysis serves as a benchmark in our simulation exercises.

4.3 Elasticities

In this section we calculate aggregate wage elasticities of the two selected models. We define the wage elasticity of labour supply as the percentage change in total working hours subsequent to a 1% increase of the before tax wage. The tax system is left unaffected.

To compute elasticities we follow the approach suggested in Duncan and Weeks (1998). This approach applies re-sampling methods to generate probabilistic estimates of employment transitions. We start with a calibration step to place individuals in their (pre-reform) observed discrete hours level. That is for each household we draw a set of pseudo-residuals $\widehat{\epsilon}_{ij}$ $(j = 1, ..., J)$ together with individual unobserved heterogeneity terms and add them to the measured utility in each of the hours points. If this results in the observed labour supply being the optimal choice for the individual, the draw is accepted; otherwise another draw is made and checked. The optimal choices after the wage increase are then predicted by the estimated deterministic model plus the stochastic elements of the model derived in the calibration step. This procedure is repeated 100 times. In this way probabilities of being in each of the discrete hours points can be obtained and expected average working hours after the wage increase can be derived.

We calculated own wage elasticities as well as elasticities with respect to the husband's wage rate. Since apart from female earnings the husband's earnings reflect the bulk of family income these can be interpreted as other income elasticities. The calculated elasticities take full account of the impact of the wage rate on the participation decision. This is because utility of working increases with the wage rate while the utility of not working is not affected by the wage rate. The results for all women as well as separately for the low and high educated are displayed in table 6. The calculated figures seem to be in line with other studies see for example Bargain (2004) or Gerfin (1993). Studies which allowed for measurement errors in the wage equation generally got larger elasticities (see Van Soest, Das, and Gong (2002)). The effects on participation presented in Table 6 are the changes in percentage points if husband's or own wage rates increase by 1%. The cross wage impact on participation is not clear. In the case of the own wage increase more than half of the elasticity is due to an effect on participation.[20]

From a policy point of view the labour supply behaviour of low educated women is of particular interest since their participation rates are much lower. According to our findings the labour supply for the low educated women seems to be somewhat more sensitive for own wages than for high educated women. Contrary the labour supply for the high educated seems to be more sensitive for husband's wage. The two models give similar results.

4.4 Labour Supply Response to a Tax Reform

In Switzerland married couples are currently taxed jointly. Given the progressive tax system this implies very high marginal tax rates for second earners and in general discrimination in favour of de-facto couples.[21] The problem is alleviated to some extent by lower tax rates for couples but there is still a marriage penalty with respect to income taxation. Of particular interest in our case are the disincentive effects of the current tax system on second earners (in most cases the woman). To asses the magnitude of these effects we simulate the female labour supply responses to the introduction of an individual tax system. The deductions are assumed to be the same as in the current system.

The way in which the effects of the reforms are predicted is very similar to the method of computing the elasticities in section 4.3. We first place individuals in their (pre-reform) observed discrete hours level using the actual tax rules and the random draws. We then predict optimal after reform choices using the tax rules according to the proposed reform and the stochastic elements of the model drawn in the first step. To take into account the probabilistic nature of the state transitions at the individual level the procedure is repeated 100 times. We assume that before tax wage rates are not changed by the reforms. In other words general equilibrium effects are not taken into account. The calculated effects are first order effects which could principally serve as inputs for a general equilibrium model.

Results are displayed in table 7-9. Table 9 displays the effects of the reform on average working hours and participation. Both models predict an increase in the participation rate and in average hours worked. The effects seem to be much stronger for the low educated than for high educated woman. Table 7-8 show transition frequencies for model S3 and model SNP respectively. The net proportion of women who would potentially take up a part-time job is approximately 1.36% and 1.55% respectively. 0.67% (0.49%) would move from non-work to full-employment and 0.2% (0.4%) would move from part-time to full employment. Disincentive effects are negligible. Again the two models give similar results.

[20] The elasticities of participation can be obtained by dividing the change in participation (in % points) by the predicted participation rates.

[21] In the case of single earner couples the discrimination is in favour of married couples.

5 Conclusions

This paper analyzes different modeling strategies for discrete choice labour supply models. The main result suggests that care should be taken when using very general functional forms of preferences in discrete choice labour supply analytics.

We compared four modeling strategies used in recent studies: a structural model with fixed costs of work and random heterogeneity, a model with a nonparametric specification of the direct utility function, a model which allows parameters to be fully alternative specific and a model that allows for price and income dependent preferences.

Some of the estimated parameters of the structural and the structural nonparametric model are directly interpretable. However as the implausible high coefficients of the fixed cost variables indicate the interpretation should be done very cautiously. What these coefficients exactly measure is unclear. Apart from fixed cost it could also be job search disutility, distaste of work or a mixture of all these. If available it seems preferable to use sample information about child care costs to account for fixed costs instead of using a shadow equation as we had to. Another alternative would be to allow the taste shifters to enter utility in a less restrictive way so that the utility function would represent preferences in which fixed costs are already captured.

A series of likelihood ratio tests show that the restrictions made in the structural model are clearly rejected. In other words the structural nonparametric model as well as the unconstrained and the non-standard models are statistically superior to the structural model. Moreover estimation of the model with price and income dependent preferences lead to a clear rejection of the standard or unitary models. However the unconstrained and non-standard model cause a significant part of the sample to not respect the only coherency restriction we really need for economic interpretation of the models and meaningful policy simulation: positive monotonicity in income. In the case of the non-standard model only 52.9% of the supply choices exhibit positive marginal utility of income. Improving results by a restricted estimation seems not to make sense here.

Overall the structural nonparametric model performs best. The model fits the data well and all supply choices exhibit positive marginal utility of income. In addition from an intuitive point of view it is not obvious why the effect of income should differ over the supply choices as in the unconstrained and non-standard model. Furthermore it is not clear if these effects remain constant after the introduction of a tax reform.

The simulation of female labour supply responses to an introduction of individual taxation in Switzerland with the structural and the structural nonparametric model give clear results. Depending on the model the effects on the participation rate and average hours worked are 4.93% (3.93) and 2.03% points (2.05) respectively. The effects seem to be much stronger for the low educated than for high educated woman.

Future research needs to go a step further and has to address the introduction of more general frameworks like the collective or the life cycle approach.

Another general approach we did not mention in our paper would consist of the disaggregation of what we call leisure into a number of different categories of time allocation (Apps and Rees 1996). The few work which has been done in these fields has shown that ignoring these approaches can lead to biased labour supply estimates. Extending the discrete choice framework in these directions could therefore be a promising direction of future research.

References

Apps, P. F., and R. Rees (1996): "Labour Supply, Household Production and Intra-family Welfare Distribution," *Journal of Public Economics*, 60(2), 199–219.

Bargain, O. (2004): "On modeling household labor supply with taxation," DELTA Working Papers 2004-14, DELTA (Ecole normale suprieure).

Blundell, R., A. Duncan, J. McCrae, and C. Meghir (2000): "Evaluating In-Work Benefit Reform: The Working Families Tax Credit in the U.K," JCPR Working Papers 160, Northwestern University/University of Chicago Joint Center for Poverty Research.

Brownstone, D., and K. Train (1998): "Forecasting new product penetration with flexible substitution patterns," *Journal of Econometrics*, 89(1), 109–129.

Callan, T., and A. V. Soest (1996): "Family Labor Supply und Taxes in Irelnad," Discussion paper, Tilburg University.

Chiappori, P. A. (1988): "Rational Household Labor Supply," *Econometrica*, 56(1), 63–90.

Creedy, J., and A. Duncan (2002): "Behavioural Microsimulation with Labour Supply Responses," *Journal of Economic Surveys*, 16(1), 1–39.

Dickens, W. T., and S. J. Lundberg (1993): "Hours Restrictions and Labor Supply," *International Economic Review*, 34(1), 169–92.

Duncan, A., and M. N. Harris (2002): "Simulating the Behavioural Effects of Welfare Reforms among Sole Parents in Australia," *Economic Record*, 78(242), 264–76.

Duncan, A., and M. Weeks (1998): "Simulating Transitions Using Discrete Choice Models," *Proceedings of the American Statistical Association*, 106, 151–156.

Gerfin, M. (1993): "A Simultaneous Discrete Choice Model of Labor Supply and Wages for Married Women in Switzerland," *Empirical Economics*, 18(2), 337–56.

Keane, M., and R. Moffitt (1998): "A Structural Model of Multiple Welfare Program Participation and Labor Supply," *International Economic Review*, 39(3), 553–89.

MaCurdy, T., D. Green, and H. J. Paarsch (1990): "Assessing Empirical Approaches for Analyzing Taxes and Labor Supply," *Journal of Human Resources*, 25(3), 415–90.

McFadden, D., and K. Train (2000): "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15(5), 447–70.

Pollak, R. A. (1977): "Price Dependent Preferences," *American Economic Review*, 67(2), 64–75.

Revelt, D., and K. Train (1998): "Mixed Logit With Repeated Choices: Households' Choices Of Appliance Efficiency Level," *The Review of Economics and Statistics*, 80(4), 647–657.

Train, K. E. (2003): *Discrete Choice Methods with Simulation*. Cambridge University Press.

Van Soest, A. (1995): "Structural Models of Family Labor Supply: A Discrete Choice Approach," *Journal of Human Resources*, 30(1), 63–88.

Van Soest, A., M. Das, and X. Gong (2002): "A structural labour supply model with flexible preferences," *Journal of Econometrics*, 107(1-2), 345–374.

Van Soest, A., A. Kapteyn, and P. Kooreman (1993): "Coherency and Regularity of Demand Systems with Equality and Inequality Constraints," *Journal of Econometrics*, 57(1-3), 161–88.

Table 1: Descriptive statistics

Variable	Women	Men
Participation rate	0.656	0.998
Hours of work (all)	17.207	41.33
Hours of work (H>0)	26.232	41.384
Gross wage rate (all)*	29.694	38.664
Gross wage rate (H> 0)	30.556	38.668
Age	36.825	39.272
High education	0.097	0.333
Low education	0.085	0.039
Net household income (per month)	7801.89	
Number of children	1.133	
Number of preschool children	0.571	
Number of schoolaged children	0.384	
Number of selected households	2305	

* Includes predicted wages for non-workers

Figure 1: Predicted and observed wage distribution

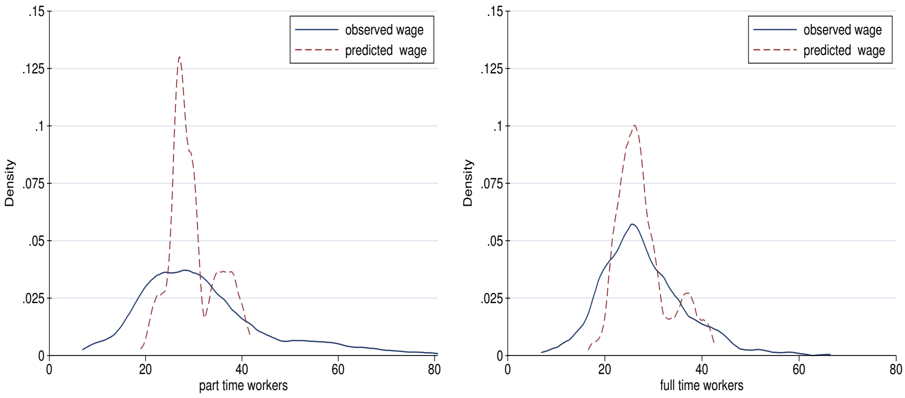

Figure 2: Distribution of female working hours

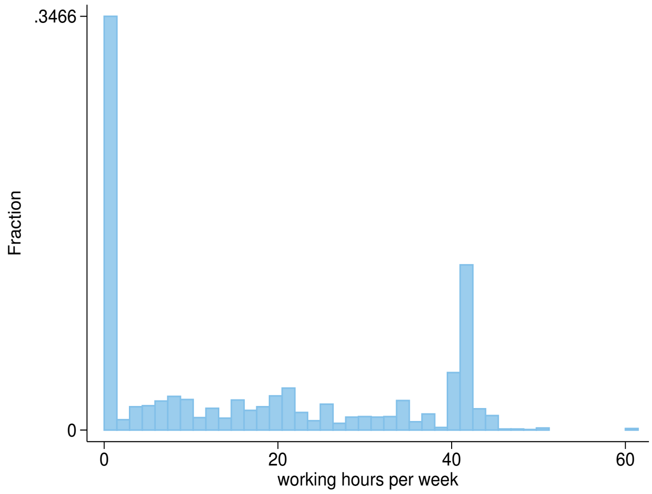

Figure 3: Indifference curves benchmark individual, model S3

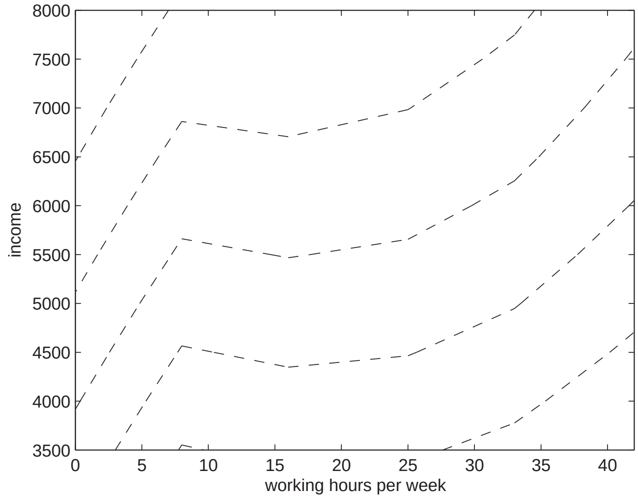

Table 2: Estimation results, model S1, S2 and S3

	S1		S2		S3	
Variable	Coeff.	(Std. Err.)	Coeff.	(Std. Err.)	Coeff.	(Std. Err.)
income^2	-0.3243*	(0.1577)	-0.4264**	(0.1246)	-0.4168**	(0.1138)
hours^2	0.1203**	(0.0176)	0.0005	(0.0243)	-0.1674**	(0.0344)
hours $\times$ income	-0.0409	(0.0788)	-0.0545	(0.0502)	-0.1350**	(0.0410)
income	3.8081**	(0.6689)	4.0415**	(0.5734)	3.4485**	(0.5225)
hours	-0.6839**	(0.1570)	0.0482	(0.1632)	1.2882**	(0.2004)
$\times$ age$-$40	-0.0312**	(0.0022)	-0.0308**	(0.0022)	-0.0366**	(0.0029)
$\times$ (age$-$40)^2	0.0002	(0.0002)	0.0002	(0.0002)	0.0003	(0.0002)
$\times$ preschool children	-0.8381**	(0.0378)	-0.8720**	(0.0472)	-0.9299**	(0.0841)
$\times$ schoolaged children	-0.4313**	(0.0327)	-0.6049**	(0.0444)	-0.8888**	(0.0823)
$\times$ high educated	0.3716**	(0.0632)	0.3853**	(0.0651)	0.4459**	(0.0770)
fixed cost/4000			0.4620**	(0.0798)	1.1339**	(0.0722)
preschool children			-0.0861*	(0.0355)	-0.0306	(0.1416)
schoolaged children			-0.2741**	(0.0487)	-0.5724**	(0.0760)
σ_y					3.1328**	(0.4386)
N	2305		2305		2305	
Log-likelihood	-3127.64		-3087.25		-3080.43	

Significance levels : † : 10% $*$: 5% $**$: 1%

Table 3: Estimation Results, Model SNP

Variable	Coeff.	(Std. Err.)
income^5	-0.0163	(0.0200)
$\text{income}^4 \times \text{hours}$	-0.0111	(0.0214)
$\text{income}^3 \times \text{hours}^2$	0.0031	(0.0345)
$\text{income}^2 \times \text{hours}^3$	0.0544	(0.0438)
$\text{income} \times \text{hours}^4$	0.1426*	(0.0598)
hours^5	0.1627*	(0.0822)
income^4	0.1007	(0.1588)
$\text{income}^3 \times \text{hours}$	-0.0349	(0.1742)
$\text{income}^2 \times \text{hours}^2$	-0.4016	(0.3265)
$\text{income} \times \text{hours}^3$	-1.3845*	(0.5434)
hours^4	-1.8678*	(0.8973)
income^3	0.0427	(0.3601)
$\text{income}^2 \times \text{hours}$	1.0363	(0.7652)
$\text{income} \times \text{hours}^2$	4.4331*	(1.7332)
hours^3	8.1155*	(3.5876)
income^2	-1.3189*	(0.6565)
hours^2	-16.7405*	(6.6290)
income $\times$ hours	-5.5376*	(2.31059
income	5.1526**	(1.9258)
hours	16.2452**	(5.6817)
$\times$ age$-$40	-0.0298**	(0.0022)
$\times$ $\text{age}^2 - 40$	0.0001	(0.0002)
$\times$ preschool children	-0.8926**	(0.0485)
$\times$ schoolaged children	-0.6339**	(0.0501)
$\times$ high educated	0.3897**	(0.0661)
fixedcost/4000	2.0901*	(0.9056)
preschool children	-0.1002**	(0.0348)
schoolaged children	-0.2456**	(0.0487)
N	2305	
Log-likelihood	-3030.80	

Significance levels : † : 10% $*$: 5% $**$: 1%

Table 4: Average Predicted Probabilities

choice	actual	S1	S2	S3	SNP	U	G
0	0.344	0.332	0.348	0.354	0.344	0.344	0.344
8	0.119	0.152	0.126	0.117	0.119	0.119	0.119
16	0.116	0.099	0.097	0.097	0.117	0.116	0.116
25	0.118	0.090	0.100	0.102	0.119	0.118	0.118
33	0.081	0.115	0.128	0.129	0.081	0.081	0.081
42	0.222	0.212	0.202	0.200	0.220	0.222	0.222
Pseudo R2		0.243	0.252	0.254	0.266	0.296	0.411

Table 5: Tests of Restrictions

mod.	log L	coeff.	vs mod.	log L	coeff.	df	LR	chi2(1%)
S3	-3127.64	14	SNP	-3030.80	28	14	99.26	29.14
S3	-3127.64	14	U	-2909.42	137	123	436.44	162.4
S3	-3127.64	14	UR	-2910.01	137	123	435.26	162.4
U	-2909.42	137	G	-2432.35	275	138	954.16	179.56

Table 6: Changes in participation and elasticities of hours worked

		husband's wage		own wage	
model		change in participation (in %-points)	elasticity of hrs worked (in %)	change in participation (in %-points)	elasticity of hrs worked (in%)
S3					
	all	0.037	-0.078	0.151	0.309
	low educated	0.078	0.096	0.216	0.734
	high educated	0.009	-0.094	0.027	0.04
SNP					
	all	-0.02	-0.208	0.133	0.237
	low educated	0.036	-0.202	0.165	0.382
	high educated	-0.067	-0.226	0.117	0.236

Table 7: Responses to the introduction of individual taxation, model S3

pre-reform	post-reform						marginal freq
	out of work		part time		full time		
out of work	32.36	(94.06)	1.37	(3.98)	0.67	(1.96)	34.4
part time	0.01	(0.02)	42.93	(98.84)	0.5	(1.14)	43.43
full time	0	(0)	0.03	(0.14)	22.14	(99.86)	22.17
marginal freq.	32.37		44.32		23.31		100

Table 8: Responses to the introduction of individual taxation, model SNP

pre-reform	post-reform						marginal freq
	out of work		part time		full time		
out of work	32.35	(94.05)	1.55	(4.52)	0.49	(1.43)	34.4
part time	0	(0)	43.02	(99.06)	0.41	(0.94)	43.43
full time	0	(0)	0.01	(0.06)	22.16	(99.94)	22.17
marginal freq.	32.36		44.58		23.06		100

Table 9: Policy effects, individual tax system and splitting model

	individual tax system	
model	change in participation (in %-points)	change in hrs worked (in %)
S3		
all	2.03	4.932
low educated	3.293	11.664
high educated	0.3	1.022
SNP		
all	2.045	3.918
low educated	3.098	8.883
high educated	0.924	1.965